Soc Psychiatry Psychiatr Epidemiol (1996) 31: 89–98

ORIGINAL PAPER

T. S. Brugha · F. Lindsay

Quality of mental health service care: the forgotten pathway from process to outcome

Accepted: 19 April 1994

Abstract The validity of the concept of outcome depends on a relationship between routine treatment and later health status. Outcome evaluations and audits are very rare in psychiatry. A substantial expansion in epidemiologically based, naturalistic, observational, process-outcome data collection in routine psychiatric practice is essential in order to identify treatment allocation biases and other reasons for unexpected outcomes. Identified causes of undertreatment should lead to locally agreed detailed clinical guidelines. Experimental evaluation should take place in routine clinical practice settings, with change in both process and outcome as the objective. Ultimately, the results of both experimental and observational outcome studies on representative service users should converge, permitting outcomes to be the ultimate arbitrator of quality.

There should be no need to justify an article on quality in a series of papers on mental health service evaluation. What does the world of scientific evaluative research have to offer? The title of this article reflects two questions about medical and psychiatric care:

1. Is care being implemented according to good practice criteria?
2. Does it work?

The first of these two questions takes account of relational aspects (especially doctor patient communication), environmental aspects (for example accessibility), and the technical aspects of the process (provision of care that is most likely to lead to a better outcome; Donabedian 1989). Whether it is worth asking may depend on the answer to the second question:

T. S. Brugha (✉) · F. Lindsay
University of Leicester, Department of Psychiatry, Clinical Sciences Building, Leicester Royal Infirmary, PO Box 65, Leicester LE2 7LX, UK

Does properly implemented *care* lead to better subsequent health *status* (and functioning, satisfaction) and if not, which (combination?) of the two should we measure and rely upon: the *process* of care? or health status *subsequent* to care provision? or a *combination* of process and health status, even when they appear to be unrelated?

The question of whether to monitor process or outcome is a major problem in quality assurance (Fauman 1990). The term health *status* is emphasised at this point rather than that of *outcome:* Donabedian (1992) defines outcomes as the *states* or *conditions attributable to antecedent health care.* If we accept this definition of outcome then it follows logically that we can only use this term when we are able to demonstrate that status is significantly associated with antecedent care. Within the field of psychiatry, outcomes might appear now to be at centre stage, having previously been off a stage dominated by structural measures of input and process measures (Jenkins 1990).

Medical statistics and epidemiological methods are concerned with two, related inseparable aims: estimation and uncertainty. If it seems to work, how little and how much does it work in this population? This can mean reducing complex sets of observational data to relatively simple, general statements that are an accurate representation of those data, whilst allowing for measurement error and *chance,* and in randomised studies (experiments) estimating the plausible range of the effect of a treatment by means of the *effect size* and *confidence interval* (Everitt 1989).

Review methodology

In preparing this article we reviewed the literature on quality of care in medicine and psychiatry and the published literature on audit in psychiatry, which is the major area of application of this topic in service settings. A literature search was conducted by searching for articles (in PSYCHLIT and MEDLINE), textbooks, cited chapters and cross references covering historical and definitional aspects, methods of assessment and examples of their use in the field

of psychiatry. A selection of relevant articles from the appropriate areas of medical statistics and epidemiology was included also.

Our first aim was to try to reach a conclusion concerning the relative benefits and feasibility of quality assessment relying on aspects of process assessment and that based on outcomes. Our second aim, also difficult to achieve, was to try to structure the evidence in the literature in a cyclical fashion as recommended in quality assurance and audit programmes: beginning with definitional issues, the establishment of quality standards, guidelines and policies; measurement and assessment issues; interpretation and appraisal as in peer group audit activities; finally, implementation strategies, including the final stage of the audit cycle and including the experimental evaluation of clinical guidelines in routine practice.

Definition and history

In 1910, Codman proposed the "end result idea" in which "every hospital should follow *every* patient it treats, long enough to determine whether or not the treatment has been successful, and then to inquire 'if not, why not?' with a view to preventing a similar failure in the future"; in essence it was equivalent to monitoring "outcomes" (Donabedian 1989). Codman suggested concurrent assessment of care and its consequences, with the occurrence of adverse outcomes being the only occasion for "process" assessment. In order to establish the relation between care and its results, observations were needed on the causes for not attaining perfection. Codman believed that the end result was the only true product of health care and the major purpose of the end result system was to bring improvements in health care (Donabedian 1989).

Quality of care is defined as the level of performance or accomplishment that characterises the health care provided (Last 1988). Structure refers to manpower, facilities, resources, numbers and qualifications of professionals, characteristics of administrative organisations and physical facilities (Tugwell 1979). Process refers to technical (investigations, physiological monitoring and treatment prescribed; diagnostic and therapeutic procedures) or interpersonal (patient education) styles (Tugwell 1979). Donabedian's (1992) definition of outcomes as the states or conditions attributable to antecedent health care is not uncontroversial. Outcome can refer to death or disability rate, disease (cure or not), effect on patient health and satisfaction (Tugwell 1979) and discomfort, social and psychological well-being.

Quality assurance and audit

Medical audit has been defined as "the systematic, critical analysis of the quality of medical care, including the procedures used for diagnosis and treatment, the use of resources and the resulting outcome and quality of life for the patient" (Department of Health 1989). In what way does audit relate to our basic question about the relationship between process and outcome? Audit is inclined to be insensitive to outcome, but sensitive to structure and process (Holman 1989). The government suggests that every consultant should be involved in a form of medical audit agreed between management and the profession locally, that it is now a contractual obligation. It is a condition for the training of junior staff; without it hospitals should not be accredited for higher specialist training (Department of Health 1989).

The aim of audit is to produce change but only if it extends to other health care workers and managers (Moss and Smith 1991). Quality assessment refers to the determination of the degree of quality of care and quality assurance refers to all measures used to protect, maintain and improve the quality of care (Donabedian 1992). Quality assurance implies a good quality service achieved at minimum expenditure, but in health care this means any procedure(s) improving quality of care (Jacyna 1992). Audit is about continuing improvement. Construction of an audit involved adopting a standard, defining an indicator, setting a target and defining the monitoring method (DeLacey 1992). The sequence of separate activities linked to and from the "audit cycle" loop should include stages of observation, comparison and action taking (Robinson 1991). The audit cycle must be completed if it is to be beneficial, that is to improve patient care (Hatton and Renvoize 1991; Moss and Smith 1991; McClelland 1992). Steps must be charted and measured. Identification of what improvements can be made should be followed by further assessment once improvements are instituted (Feldman 1992). The operational definitions of quality assurance all have the feedback cycle in common (McClelland 1992).

Quality standards and practice guidelines

Policy aspects

In the United Kingdom, the Audit Commission (1992) has a statutory duty to promote economy, efficiency and effectiveness in bodies that it audits, which since 1990 has included the National Health Service (NHS). Its role is to prioritise the patient perspective, community care and joint audits, and to develop tools for direct use, quality exchange, accreditation and league tables.

Standards in psychiatry

Standards of care will depend increasingly on regularly updated overviews and meta-analyses of evidence of the efficacy of psychiatric treatments and related interventions (Wing 1992; Depression Guideline Panel 1993). The diversity of professional providers in psychiatry has also complicated the development of standards, classifications of intervention problem groupings, and thus methods for monitoring the quality of the process of care (Wells and Brook 1988). Both national governments and agencies and the World Health Organisation have promoted standards.

The *Health Advisory Service* (HAS), The *Mental Health Act Commission Biennial Reports* (MHAC 1993), the MHAC Second Opinion system, the Mental Health Review Tribunals and the Approval Exercise of the *Royal College of Psychiatrists* are all examples of formal institutional audit (Garden et al. 1989). The introduction throughout the NHS of the *Care Programme Approach* (Royal College of Psychiatrists 1991) has been initiated through similar statutory procedures.

The Royal Australian and New Zealand College of Psychiatrists set up the "Quality Assurance in Aspects of Psychiatric Practice" project (Holman 1989), which has resulted in the development of possibly the first ever treatment recommendations for depressive disorders (Quality Assurance Project 1983). It is concerned with more than just audit; a series of treatment outlines for major conditions were developed as a basis for peer review and research. Holman (1989) has suggested that a clinical focus should be maintained in audit especially by the use of care plans and established guidelines similar to the Australian Quality Assurance Programme. More precise treatment guidelines are also being developed elsewhere [Depression Guideline Panel (1993)], making use of diagnostic and treatment decision trees and algorithms. The development of more precise guidelines will facilitate quantitative audits of undertreatment leading to the prioritisation of practice altering evaluation projects (Brugha, in press).

Quality measurement

Quality is judged by individual professionals comparing it with a standard; but it may be perceived differently by users. Donabedian (1966) consideres the sources and methods of obtaining information: sampling and selection, clinical research including the limitations of direct observation especially in general practice, measurement standards (empirical and normative), measurement scales and reliability, bias and validity.

The classic work of Donabedian is fraught with the problems of using each dimension (e.g. structure) in isolation (Turner 1989). Turner (1989) suggests other quality of care perspectives. First, patient perceptions, that is patients may judge quality more by how they are treated than by the health outcome. Second, adherence to standards, that is from industry, a multidimensional approach, but with the focus still on outcome. Monitoring quality in medical practice has come to be synonymous with the growing practice of audit.

There has been a long debate on the right way to measure the quality of care, whether to use process or outcome criteria (Ierodiakonou and Vandenbroucke 1993). These workers have argued that the ultimate judgement of quality rests on the evaluation of process [believed by the ancient (Greek) philosophers and modern theoreticians of quality assurance; Ierodiakonou and Vandenbroucke 1993]. Some administrators are convinced that quality of performance should be measured according to what they assert to be outcome criteria, e.g. mortality, but there are dangers involved in ranking (Ierodiakonou and Vandenbroucke 1993). To use outcome as a quality measure, continuous evaluation of all individual patient characteristics is needed, which is a gigantic and perhaps unrealistic research effort (Ierodiakonou and Vandenbroucke 1993).

Tugwell (1979) clearly advocates process-based approaches in a quote from Cochrane: "the core of quality of medical care is the extent to which scientifically proven effective methods of treatment are properly applied to patients who can benefit from them." A strong case for a process-driven quality of care strategy has been made more recently by Micossi et al. (1993).

Evaluation of structure and input

The considerable emphasis on structure particularly in governmental policies, for example on deinstitutionalisation, has not been accompanied by published measures of structure. Readers may be acquainted with standard forms used in official institutional inspections and educational programme accreditation exercises (Garden et al. 1989); however, their status as measurement tools remains uncertain. The World Health Organisation (Janca and Chandrashekar 1993) has published details of six instruments designed to be used in quality assurance assessments: these consist of national assessments of mental health policy, mental health programmes, outpatient mental health facilities, and within a given setting, assessments of primary health care facilities and residential facilities for the elderly mentally ill. The former cover such matters as decentralisation, equity, community participation; the latter cover such matters as cleanliness, privacy, water and food. Both types of measure cover such matters as staffing, physical environment, interaction with families and the community. Six language versions are available or are in preparation.

Measurement of process

Micossi et al. (1993) have argued that since an outcomes-based approach is either impracticable (randomisation is rarely feasible) and usually unreliable (due to unknown imbalances in treatment allocations) a *profiles of care* approach is preferable, being driven by the symptoms presented by patients when first seen by a physician, which determines the resources utilised and the costs incurred. Profiles of care represent blocks of symptoms and/or intermediate diagnoses that are associated with corresponding objectives and procedures. Quality control can therefore be based on the comparison between observed and expected actions.

Does psychiatry have any examples of such process-based approaches to quality assessment? Shepherd (1988) argues that the most systematic approximation

to a process method of quality assessment in the field of psychiatry is the "Needs for Care Assessment" originally described by Brewin et al. (1987). This is based on an individualised assessment of clinical and social problems or deficits in functioning, linked with a schedule that prescribes appropriate actions or "forms of care" for the defined problems; progress with its use has been discussed more recently (Brewin and Wing 1993). According to this more recent report, several groups of researchers have been able to achieve an acceptable level of reliability in the use of this method, although it relies on the use of judgements both of what constitutes potentially worthwhile care and of whether realistic attempts have been made to provide it.

Process measures must be considered in the context of agreed standards of treatment, but it has been argued that there is a problem in the variety of approaches and modalities (Turner 1989). Tugwell (1979) proposes methodological criteria to assist process measurement. First, Tugwell proposes criterion validity: a statistical association is needed between process and outcome measures. His review of the literature shows few correlations in process-outcome studies. Methodological reasons have been suggested, for example sample size, inappropriate sampling and inappropriate measures. Second, he proposes clinical credibility of the process criteria with health professionals: decreased credibility would occur if items are unlikely to influence management. Third, he proposes accuracy: a measure must reflect actual clinical process. However, physician questionnaires may lead to an alteration in their usual clinical behaviour. Fourth, he proposes comprehensiveness: items must include all important aspects of the process of care. For example, patient education is often omitted. Fifth, he proposes sensitivity to differences between practices and sensitivity to improvements or deteriorations over time. Sixth, he proposes amenability to index construction: the results should enable statistical analysis. Seventh, he proposes feasibility and cost: measurement must be simple and acceptable. This could be achieved in a number of different ways, for example the use of record review, direct observation and the use of physician and patient questionnaires.

Williamson (1971) has developed a strategy for *process and outcome assessment.* The strategy is based on factors likely to have the greatest probability of effecting significant improvement in health status of a target population. The four elements of the strategy are diagnostic, therapeutic, process and outcomes. The study involves the development of outcome criteria; to determine whether process study is required, a comparison is made with the outcomes achieved. As a result of the process study, the direction and priorities for action to improve outcomes should be established. The strategy can be seen to enhance educational effectiveness.

Brook (1977) has questioned the validity of process criteria; only "technical" not "humanitarian" aspects of care were being measured. Brook (1977) cited a study that reported no relation between a process and an outcome assessment of quality of care and so invalidated process audit. He suggests that one should focus on very simple process criteria. Alternatively, efforts to bypass measuring process of care and concentrate on outcome could be considered, possibly by using short-term "proximate" outcomes (Brook 1977).

Outcome assessment

Donabedian (1992) believes that outcomes are the paramount criteria of good quality; that is that they remain the ultimate validators of effectiveness and quality of medical care (Fessel and Van Brunt 1972). Donabedian (1992) has drawn up a classification of outcomes of health care and has discussed the uses of outcomes in quality assessment. For example, outcomes only permit inference (not direct assessment) about process (and structure); with the role of intercurrent factors, outcomes may be misleading indicators. Outcomes are "integrative" also, that is they are of value but need process analysis.

Outcome indicators

An indicator is a measurable variable related to facilities, treatment or outcome of care (Fauman 1990). The identification of indicators and the definition of clinical criteria is a specialised task and extra training is needed. Measurements of process and structure are only acceptable as quality indicators if they predict (outcome) functional status or patient survival (Tugwell 1979).

Jenkins (1990) has proposed a system of outcome indicators for mental health care needed for monitoring and evaluation by clinicians, District Health Authorities (DHAs) and Directors of Public Health (DPH). Jenkins (1990) considers the indicators of input, process and outcome for schizophrenia, affective psychosis, neurosis, dementia, child psychiatry, forensic psychiatry, mental handicap, disability and mortality. She regards process indicators for all illness types simply as "activity on (input indicators)". Jenkins (1990) concludes that it is more useful to measure inputs and outcomes and only use process measures when necessary to investigate shortfalls in achieving objectives.

Outcome scales are presently under development as part of the Health of the Nation target-setting strategy (Royal College of Psychiatrists Research Unit 1993, in preparation). Although these scales cover health and social functioning in the form of current status measures, their use nationally will allow variations in the distribution of health service resources to be compared with non-matching prevalence and severity rates of mental health problems in districts and regions. Their widespread introduction and storage in databases may make it possible to conduct sophisticated statistical analyses of process and outcome.

Outcome measures in psychiatry are complicated and simple quality of life measures are not available

(Roy 1991). Gath (1991) has discussed questioningly the use of consumer satisfaction, that is of the patient and family. Unsolved problems include the timing of outcome measures, since there are long periods of time involved; there are few reliable outcome measures and psychiatric patients have multiple problems. Outcome measures in psychiatry are complex over long time periods; they involve subjective feelings and are greatly dependent on the patient's involvement and motivation (Turner 1989). Turner (1989) has suggested that it is important to ask "Outcome for whom? – the patient, the patient and family?".

There are problems with the use of outcome measures; many variables, in addition to the process of care itself, may contribute to the final outcome (Tugwell 1979) so that poor outcome does not necessarily imply poor quality of care (Fauman 1989), and, arguably, good outcome does not mean that credit can be apportioned to the health care system. Therefore, risk factors or covariates must be controlled for in any analysis. Confounding is defined as the failure of a crude association to reflect properly the magnitude and direction of an exposure effect because of a different distribution of extraneous risk factors among exposed and unexposed individuals (Datta 1993). A confounder is associated with a disease (outcome) and exposure (process) factor, and is extraneous to two main variables but can distort their relation (Datta 1993). Once a strong cause effect relationship has been established, process can be monitored as a surrogate for outcome of care (Fauman 1989).

Methods of audit

Robinson (1989) outlines the various methods of audit that exist; some involve process measures and others, outcome. Case note review has been used by the Royal College of Physicians audit in the reaccreditation of training posts. Criterion-based audit is utilised in peer review. Outcome audit is the most sophisticated and valid, but has difficulties. Information-based audit involves a review of aggregated activity and financial data. Topic-based and intermediate outcomes are two other forms of audit. Hatton and Renvoize (1991) also consider the use of a random case note sample that is criterion-based and covers adverse occurrences.

In the USA, audit has been performed by local Professional Standards Review Organisations (PSROs) and the Joint Commission on Accreditation of Hospitals (JCAH). Both were found to be costly and without obvious benefit (Garden et al. 1989). The JCAH focused on diagnosis-related groups (DRG). In Canada and the Netherlands there is a legal requirement to perform quality of care or quality assurance programmes. The JCAH, later known as the Joint Commission on Accreditation of Healthcare Organisations (JCAHO), became involved in monitoring due to public demands for accountability and requirement for institutional accreditation (Fauman 1989). Their new approach involved an emphasis on clinical outcome rather than on delivery of care. Criteria could be classified in relation to structure, process of outcome, they could be implicit or explicit (specified in advance), referents (the problem or diagnosis to which criteria apply or have a normative or empirical source (derived by consensus or by empirical investigation). Indicators, tracers (broadly defined health problems) and thresholds triggering more intensive evaluation could be used in monitoring (Fauman 1989). The JCAHO planned to identify indicators (of outcome) in psychiatry and became the main driving force in the development and application of standards of quality of medical care (Fauman 1990). The JCAHO developed an audit system, the performance evaluation procedure (PEP), which was later discontinued.

There is a conflict between ensuring quality of treatment and controlling expenditures; therefore, attempts are required to link quality assurance with "cost-effectiveness analysis" (Cahn and Richman 1985). There is a distinction between quality assessment and quality assurance; quality assurance means measuring both the level of care and, when necessary, improving it (Cahn and Richman 1985). The processes of quality assurance include "medical audit" and "PEP".

Benefits and concerns

There is criticism that little attention is given to patient's desires or perceptions of treatment effects. The Department of Health (1989) acknowledges the inevitable differences between audit in the hospital, community and primary health care. For a satisfactory audit (of representative users), a suitable form of case register is required (Daly 1991). Furthermore, it is important to use doctors time efficiently (Gath 1991). In order to disseminate the explosion in output of audit reports, in March 1992 the BMJ Publication Group launched a new journal *Quality in Health Care*. The BMA and BMJ have also set up a joint working group on quality (Moss and Smith 1991). Since 1989, the *Bulletin* of the Royal College of Psychiatrists has regularly carried brief articles on audit. We have reviewed this literature in greater detail in a separate report (Brugha and Lindsay 1994), considering here only those few studies that provide clear findings about the relationship between process and later health status (i.e. outcome).

Quality of care studies in psychiatry

The influence of health care on suicide is uncertain; it has been considered by a number of writers to be an important mental health service outcome indicator (Hawton 1987; Jenkins 1990). Morgan and Priest (1991) have carried out a study following on an initiative by the Royal College of Psychiatrists; in essence it was an audit of unexpected deaths. Demographic and clinical data, including diagnosis and treatment, were collected by

means of a questionnaire completed by the responsible consultant. The results pointed to a number of possible risk factors for suicide and other unexpected deaths; the included misleading clinical improvement in the absence of corresponding alleviation of situational problems, and social alienation of the patient. The study was felt to have implications for service development, with major reductions in bed numbers planned; this method of audit would need to be evaluated.

Structure, process and outcome quality evaluations in psychiatry

Structure

Education can be considered to be a structural influence on the process of care. Rutz and his colleagues (1992) have carried out a study in which they followed up the long-term development of an educational programme for all general practitioners (GPs) on the prevention and treatment of depression. The educational programme was completed by all GPs on a Swedish island. Process and outcome measures of the quality of care, including the number of referrals, the number of emergencies, sick leave, prescription of psychotropics, inpatient care and suicide frequency, were made before and after the programme. The results of the study have indicated that the effects, strictly related in time to the educational programme, which included a lowered suicide rate, were real and not only a coincidence with local trends. They concluded that the educational programme significantly affected important areas of the health care system. They suggested that such educational programmes should be provided every 2 years. This open study represents one of the most compelling pieces of evidence that suicide can be an outcome of antecedent care, although the results should be cautiously interpreted until similar work is conducted by means of a random allocation design in a more representative setting.

Process

Previous reviewers have noted the small number of studies that have focused on the process of psychiatric care and particularly on aspects of drug treatment (Wells and Brook 1988). The Needs for Care Assessment System referred to earlier has been used in a number of studies evaluating the process of care, particularly for long-term patients (Brewin and Wing 1993). The first such study carried out on 145 long-term users of psychiatric day care showed that benzodiazepine tranquillisers and anticholinergic preparations were being used frequently without their need being reviewed by the responsible clinician; episodes of depression and anxiety disorder were sometimes untreated and psychotic symptoms were often undertreated; deficits in role skills that were being particularly neglected included self-care and literacy skills, for which remedial training or shelter was unlikely to have been offered (Brewin et al. 1988). When a similar methodology was applied to physical health problems in the same population (Brugha et al. 1989) almost half (44 %) of those with such problems had not received appropriate assessment or treatment.

In a recent report on a cohort of 119 adults with hospital-treated depression, at 3- to 6-month follow-up, 4 out of 5 of those who were still not recovered had been offered no specific treatment or no change in their previous treatment (Brugha and Bebbington 1992). Two other larger, prospectively assessed cohorts have also documented a fall-off in the use of efficacious treatments after approximately 2 months of active treatment, even in those who have not recovered (Shea et al. 1992; Rogers et al. 1993), as well as a general failure to initiate potentially effective treatments in the first place. Process evaluation has also been used in a primary care based study of quality of care (Sibley et al. 1975), in which depression was one of a list of 10 common indicator conditions evaluated according to detailed predetermined criteria. Although drug treatment was scored as adequate in about 40 % of uses, other aspects of the management of depression (frequent follow-up assessments and support) were scored as adequate in about 80 % of cases identified.

Outcome

Although there have been observational, outcome studies of community care initiatives and psychotherapy services, in general, there is very little outcome research in relation to the process of psychiatric care. In a recent overview of the strength and quality of evidence for effectiveness of treatments for neurotic, affective and functional psychotic disorders (Wing 1992), no citations were based on well-designed cohort or case-controlled analytic studies (from more than one research group). In one open study evaluating a de-institutionalisation programme, better clinical status appeared to be associated with higher costs (Beecham and Knapp 1992). Schuster (1991) has reported that outcome studies will be critical in preventing further limitations in psychiatric care and its funding. He has suggested that quality and cost should be improved by concentrating on treatment settings and who gives the treatment (i.e. process). The Medical Outcomes Study is a 2-year, prospective, observational study in which depression is one of four chronic conditions under study (Tarlov et al. 1989). In the next section we refer to a key report on the relationship between structure (fee payment), process and outcome of depression.

The structure-process-outcome paradigm

The structure-process-outcome paradigm provides information from which inferences about quality of care may be made; that is they are not attributes of quality unless they are causally related. It has been argued that, in the future, process and outcome should be measured together (Williamson 1971; Fessel and Van Brunt 1972; Wells and Brook 1988). None of the studies in which the Needs for Care Assessment System, referred to earlier, was used have examined, as an outcome validator, the relationship between these very detailed indices of quality of care process and later health status. In a recent report on a cohort of 119 adults with hospital-treated depression, outcome at 3–6 months was not explained by prior treatment with medication, even when relevant predictors of outcome such as initial severity of illness were adjusted for (Brugha et al. 1992). Although these workers were unable to find any detectable bias in the way treatments were assigned, they have argued that outcome studies of this kind can be interpreted in a misleading way because of the non-randomised, observational design employed.

To our knowledge, only one report has appeared to date on predictors of outcome (Rogers et al. 1993) from the Medical Outcomes Study referred to previously. This has shown that a poorer outcome in prepaid services, compared with fee for service financing, is apparently due to an early fall-off in antidepressant treatment in the prepaid-financed service. Arguably this study demonstrates a link between structure (payment procedures), process (drug treatment) and later functioning (outcome), but as the authors point out their non-experimental evidence cannot claim the status of proof. Indeed, the authors were reluctant to report their process-outcome findings as so few of the outpatient cases of depression had received any treatment and there was, not surprisingly, a tendency to treat more actively the more severely ill cases (Wells et al. 1992). However, their study does have the authenticity of patient population representativeness, a major advantage that cannot be confidently claimed for willing participants in randomised trials (Cross Design Synthesis 1993).

Implementing change

Glick and his colleagues (1989) discuss the reasons for disparity between the quality of the scientific base and quality of care. They outline the obstacles to quality of care. A central failing of quality assessment is that it is rarely used to change behaviour (Brook 1977). However, the shortcomings of intervention studies are the lack of internal and external validity of "outcome" measures (Moskowitz 1993).

Purchasing for quality

A criticism of recent attempts to reform the management of the NHS has been that *activity* has been the principal measure of performance: the more health care provided the better (Sheldon and Borowitz 1993). Improvement in quality will thus depend on a shift from purchasing activity to purchasing effective technology; but how is this to be achieved?

One suggestion is for purchasers to contract for evidence-based protocols (Sheldon and Borowitz 1993), but with the exception of depression, these have yet to be developed (not to mind evaluated) in psychiatry. Yet there is evidence elsewhere in medicine that some patients receive care they do not need, some are denied care they could benefit from and that these discrepancies occur not only in well-but also in poorly resourced and funded settings (Gill 1993). The way that providers organise and monitor their own activity and thus quality is therefore an important topic. In our companion article (Brugha and Lindsay 1994) we discuss management styles, such as Total Quality Management, which may have a valuable contribution to make to the task of implementing change in the future.

Experimental and independent evaluation

Although work has been carried out on the effects of different payment methods on later mental health outcomes (Rogers et al. 1993), we have not been able to obtain any evaluative evidence of the effectiveness of such organisational and management strategies in relation to psychiatric services and outcome. Clearly, if there is any prospect of their widespread acceptance, they should be the subject of experimental evaluation.

There is encouraging evidence already of the beneficial effects of clinical guidelines experimentally introduced locally into medical practice (Grimshaw and Russell 1993). These reviewers have found a number of factors that influence whether guidelines are accepted and implemented. In a limited number of experimentally evaluated studies, evidence in most cases is that when guidelines are adopted in practice, outcomes can be empirically demonstrated by an association between increased adherence to guidelines and subsequent enhanced health status. The durability of such changes in practice is not known.

None of the cited experimental studies has focused on structural or process aspects of psychiatric care, although of possible relevance is separate research showing that when different methods of fee payment are randomised, clinical outcomes are not different (Rogers et al. 1993). Thus, no studies experimentally evaluating the effects of clinical guidelines and protocols on the process and outcome of routine care have been conducted in psychiatric services although we know of unsuccessful attempts to obtain funding support for such work and although such protocols are now beginning to

become available (Depression Guideline Panel 1993). We would urge caution about the premature introduction of guidelines in purchaser/payer contracts until their benefits have been empirically tested.

Discussion

Clearly, the difficulties that we encountered in relating process and outcome do not apply to enormous and dramatic effects, such as the effect that inhumane, degrading or punishing environments and regimes clearly have on patients' quality of life. Our difficulties were to do with less substantial and obvious effects, some of them delayed over time, such as the effect of a course of antidepressants, or a series of cognitive behaviour therapy sessions on depressive or anxiety symptoms weeks and months later. It is clearly recognised that it is only for these less substantial effects, which can be difficult to detect and demonstrate in an unbiased way, that sophisticated instruments and research designs are required (NHS Management Executive 1992). Randomised designs are a fundamental part of any such strategy. Before saying anything further about experimental methods, what can be said concerning observational methods, given that most quality assurance activity will be based in some way on these? First, the case against and then the case for observational methods.

If, as some would argue, we cannot be certain that when care is not randomly assigned later status is not due to other antecedent factors that have not been considered or measured (Datta 1993), then later status cannot be relied upon. Arguably, therefore, "outcome measurements cannot be adopted as standard tools to assess the performance of healthcare facilities" (Micossi et al. 1993). For example, having identified post-treatment health status indicators (perhaps erroneously assumed to be outcomes) that are less than optimal, attention may focus logically on the supposedly antecedent factors of structure and process in that order. Whether this is a good or a bad thing depends also on the appropriateness of the targets and the effects on the health care system of any change in focus: "setting inappropriate targets often has the effect of diverting effort from the legitimative activity of the organisation" (Lancet 1993).

In effect, in real world practice settings, if outcome cannot be relied upon then quality can only be judged by assessing the extent to which care that service users are capable of benefiting from is provided according to criterion standards. According to this argument, measurement of quality should be based on the size of the gap between observed and expected (ideal) care actions. This brings the focus back to process and the structural factors (service resources, training and organisation) that underpin care activity. If so, should these criterion standards be determined from scientifically verified evaluations of the efficacy of care actions? Unfortunately, scientifically verified evaluations, which in conventional practice means randomised trials, may not be the perfect "yardstick" for setting down such standards because "the way that patients are recruited for a randomised study can seriously impair the generalisability of results" (see Cross Design Synthesis: A new strategy for Medical Effectiveness Research, US GAO B244808, 1992).

We have gone to great pains to identify reports of prospective data on treatment outcomes in routine practice settings, but mostly to no avail. We share the widely acknowledged reservations about the reliance that can be placed on data on treatments that have not been randomly allocated. However, we are concerned that potentially useful data sets have not been analysed by means of the more advanced and rigorous methods of analysis now available (Cross Design Synthesis 1992). The data available are strikingly inadequate and incomplete; but, where available reveal an unpalatable lesson, which is that existing routine practice has, at best, extremely weak beneficial effects. We could choose to conclude that such existing "routine practice" data are consistently erroneous and that *only* randomised experiments can be relied upon. However, much of the evidence that we accept about the aetiology of disease is based on research using precisely these observational methods: are we being inconsistent if we fail to reject those findings also in their entirety?

Thus, our knowledge of "effective" treatments (again excluding major effects) is based almost exclusively on randomised experiments. These are conducted in unrepresentative ways on unrepresentative and willing subjects. Should it be surprising that, perhaps, the same "good outcomes" might not occur in routine clinical practice (Kupfer and Freedman 1986)? In routine clinical practice, diagnostic and treatment protocols are a rarity, and treatment compliance, which is probably acceptable in 50% of cases, is not directly monitored through tablet counting or drug metabolite monitoring (Wright 1993); failure to attend for non-tablet treatments (day care, psychotherapy) is unlikely to receive the same urgent attention as it does in treatment trials, unless there is a very real concern of self- or other-directed harm.

How unrealistic is it to demand further experimental evidence? Since traditional randomised clinical trials have tended to furnish data on narrow, unrepresentative subsets of the total population of those attending health services (Cross Design Synthesis 1992), should we be trying to devise new research designs beginning with the aim of minimising patient exclusions that would not occur in day-to-day clinical practice? This may mean randomising structural and process variables, that is ways of treating people, rather than randomising different treatments to each person (as in orthodox treatment trials of the kind that we also continue to need). In this article we referred to encouraging evidence that this approach can lead to improvements in the process and outcome of care (Grimshaw and Russell 1993).

In discussions of quality of psychiatric care, in what way does the process versus outcome debate apply to

the major and the commoner psychiatric disorders? First, evidence for potential effectiveness (i.e. efficacy) has been demonstrated, for the most part, in randomised-designed studies in which clinical outcome tends to be assessed over a single or brief period of time. But it is increasingly being recognised that recurrence and chronicity rather than prolonged remission characterises most of these disorders. How should clinical management protocols for depression, already referred to, increasingly define, as targets of intervention, remission maintenance, relapse prevention (Depression Guideline Panel 1993) and altered management for non-responders (Brugha, in press)? If purchasers and payers are to contract for quality based on demonstrable effectiveness, and therefore outcomes, confirming that they are getting what they are paying for, how should this be effected? Will the call for contracts based on treatment protocols (Sheldon and Borowitz 1993) be right for psychiatry (assuming that trials of protocols yet to be commissioned and completed confirm their value (i.e. effectiveness) in routine practice settings)?

So where does this now leave purchasers and providers with responsibility for *assuring quality?* Major changes in the structure, including the management of health services, could act as an ideal opportunity for experimental studies of the kind argued for in the last two paragraphs. For the present, changes in practice should follow the systematic route of adopting process protocols that reflect best clinical practice. This route may be forced upon the medical profession from purchasers and payers unless the profession itself guides its introduction (Horton 1993). But should we also make use of Codman's nineteenth century lesson (Donabedian 1989) that when the end results of health care are less than expected, that is the time to go back and ask why. The case for an outcomes-managed health service is growing on both sides of the Atlantic (Jenkins 1990; Ellwood 1988). How should this be achieved at a local level? There is little evidence to help us answer this question. As contributors to a journal of social psychiatry, it may not surprise readers that we favour exploring solutions that give serious attention to environmental and social aspects. We are less impressed with the arguments for admonishment (try harder, work harder) and more impressed with arguments for changes in the organisation of social systems (management structure). Thus, we suspect a team might do rather better at implementing outcomes management as a working group, whether locally based or at a wider but more removed level. At least this should be experimentally evaluated in relation to process and outcome measures.

When deficiencies in care are identified and localised, we would also endorse the case for a form of clinical supervision based on direct observation and feedback by a recognised expert in the field (Wells and Brook 1988). The educational effectiveness of such direct feedback teaching methods in achieving measurable enhancements in skills has been clearly demonstrated in the area of doctor patient communication and clinical assessment (Maguire et al. 1978); many recent medical graduates are already accustomed to this style of learning and would find it acceptable. Overcoming deficiencies and maintaining improvements may be crucially dependent on a shared clinical information system also. In contrast to these approaches, peer review meetings, for example of audit groups of the kind recently recommended to psychiatrists (The Royal College of Psychiatrists 1989), may be of limited effectiveness (Stocking 1992) (again, effectiveness has yet to be demonstrated).

Until the lessons of a more empirical, scientific approach, which we have attempted to face up to honestly here, begin to be more widely accepted and implemented in practice, we would answer the question we began with by recommending that, in the shorter term, the process of care should be monitored; that monitoring should be particularly pursued when associated outcomes are less than expected. In the longer term we should aim to be able to demonstrate that randomised trial proven technologies do lead to measurable improvements in outcome throughout the population of those capable of benefiting from them. Our ultimate aim for public health should be to base quality of care assessment on outcome. This will only happen through a substantial investment in medical effectiveness technologies (Cross Desing Synthesis 1992; NHS Management Executive 1992) in order to determine the effectiveness of changes in the management of services, the structure and process of care and a rational future for public health policy at local and central levels.

References

Audit Commission for Local Authorities and the National Health Service in England and Wales (1992) Minding the quality. Audit Commission, London

Beerham J, Knapp M (1992) Casting psychiatric interventions. In: Thornicroft G, Brewin CR, Wing J (eds) Measuring mental health needs. Gaskell, London, pp 163–183

Brewin CR, Wing JK (1993) The MRC Needs for Care Assessment: progress and controversies. Psychol Med 23: 837–841

Brewin CR, Wing JK, Mangen SP, Brugha TS, MacCarthy B (1987) Principles and practice of measuring needs in the long-term mentally ill: the MRC Needs for Care Assessment. Psychol Med 17: 971–981

Brewin C, Wing J, Mangen S, Brugha T, MacCarthy B, Lesage A (1988) Needs for care among the long-term mentally ill: a report from the Camberwell High Contact Survey. Psychol Med 18: 457–468

Brook RH (1977) Quality – can we measure it? N Engl J Med 296: 170–171

Brugha TS (in press) Depression undertreatment: lost cohorts, lost opportunities. Psychol Med

Brugha TS, Bebbington PE (1992) The undertreatment of depression. Eur Arch Psychiatry Clin Neurosci 242: 103–108

Brugha TS, Lindsay F (1994) Quality outcomes in Psychiatry – background and policy development. A report for Leicestershire Mental health Services

Brugha TS, Wing JK, Smith B (1989) Physical health of the long-term mentally ill in the community: is there unmet need? Br J Psychiatry 155: 777–781

Brugha TS, Bebbington PE, MacCarthy B, Sturt E, Wykes T (1992) Antidepressives may not work in practice: a naturalistic prospective survey. Acta Psychiatr Scand 86: 5–11

Cahn C, Richman A (1985) Quality assurance in psychiatry. Can J Psychiatry 30: 148–152
Cross Design Synthesis: a new strategy for medical effectiveness research (1992) Report no. B244808 Washington D. C., US GAO
Daly OE (1991) Reading about ... medical audit. Psychiatr Bull 15: 209–210
Datta M (1993) You cannot exclude the explanation you have not considered. Lancet 342: 345–347
DeLacey G (1992) What is audit? Why should we be doing it? Hosp Update June: 458–466
Department of Health (1989) Working for patients. Medical audit. Working paper 6. HMSO, London
Depression Guideline Panel (1993) Depression in primary care, vol 2. Treatment of major depression. Clinical practice guideline number 5. Department of Health and Human Services, Public Health Service, Agency for health Care Policy and Research. AHCPR publication no. 93-0551, Rockville M. D.
Donabedian A (1966) Evaluating the quality of medical care. Milbank Memorial Fund Q 44: 166–206
Donabedian A (1989) The end results of health care: Ernest Codman' contribution to quality assessment and beyond. Milbank Q 67: 233–256
Donabedian A (1992) The role of outcomes in quality assessment and assurance. Qual Rev Bull 18: 356–360
Ellwood PM (1988) Outcomes management: a technology of patient experience. N Engl J Med 318: 1549–1556
Everitt BS (1989) Statistical methods for medical investigations. Oxford University Press, New York
Fauman MA (1989) Quality assurance monitoring in psychiatry. Am J Psychiatry 146: 1121–1130
Fauman MA (1990) Monitoring the quality of psychiatric care. In: Soreff S, Uttermohlen D (eds) Psychiatr Clin North Am 13: 73–88
Feldman MM (1992) Audit in psychotherapy: the concept of Kaizen. Psychiatr Bull 16: 334–336
Fessel WJ, Van Brunt EE (1972) Assessing quality of care from the medical record. N Engl J Med 286: 134–138
Garden G, Oyebode F, Cumella S (1989) Audit in psychiatry. Psychiatr Bull 13: 278–281
Gath A (1991) Audit. Psychiatr Bull 15: 23–25
Gill M (1993) Purchasing for quality: still in the starting blocks? Quality Health Care 2: 179–182
Glick ID, Showstack JA, Cohen C, Klar HM (1989) Between patient and doctor. Improving the quality of care for serious mental illness. Bull Menninger Clin 53: 193–202
Grimshaw JM, Russell IT (1993) Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet 342: 1317–1322
Hatton P, Renvoize EB (1991) Psychiatric audit. Psychiatr Bull 15: 550–551
Hawton K (1987) Assessment of suicide risk. Br J Psychiatry 150: 145–153
Holman C (1989) Medical audit in psychiatry. Psychiatr Bull 13: 281–284
Horton R (1993) Data-proof practice. Lancet 342: 1499–1500
Ierodiakonou K, Vandenbroucke JP (1993) Medicine as a stochastic art. Lancet 347: 542–548
Jacyna MR (1992) Audit assesses quality: but what is quality? A clinician's view. Hosp Update November: 822–824
Janca A, Chandrashekar CR (1993) Catelogue of assessment instruments used in the studies co-ordinated by the WHO Mental Health Programme. WHO/MNH/92.5, Geneva
Jenkins R (1990) Towards a system of outcome indicators for mental health care. Br J Psychiatry 177: 500–514
Kupfer DJ, Freeman DX (1986) Treatment for depression. 'Standard' clinical practice as an unexamined topic. Arch Gen Psychiatry 43: 509–511
Lancet (editorial) (1993) Dicing with death rates. Lancet 341: 1183–1184
Last JM (ed) (1988) A dictionary of epidemiology. Oxford University Press, New York
Maguire P, Roe P, Goldberg D, Jones S, Hyde C, O'Dowd T (1978) The value of feedback in teaching interviewing skills to medical students. Psychol Med 8: 695–704
Mental Health Act Commission (1993) Mental Health Act Commission; fifth biennial report 1991–1993. HMSO, London
Micossi P, Carbone M, Stancanelli G, Fortino A (1993) Measuring products of health care systems. Lancet 341: 1566–1567
McClelland R (1992) The quality issue. Psychiatr Bull 16: 411–413
Morgan HG, Priest P (1991) Suicide and other unexpected deaths among psychiatric in-patients. Br J Psychiatry 158: 368–374
Moskowitz JM (1993) Why reports of outcome evaluations are often biased or uninterpretable. Eval Program Plann 16: 1–9
Moss F, Smith R (1991) From audit to quality and beyond. BMJ 303: 199–200
NHS Management Executive (1992) Assessing the effects of health technologies. Department of Health, London
Quality Assurance Project (1983) A treatment outline for depressive disorders. Aust N Z J Psychiatry 17: 129–146
Robinson M (1991) Medical audit: basic principles and current methods. Psychiatr Bull 15: 21–23
Rogers WH, Wells KB, Meredith KB, Sturm R, Burnham A (1993) Outcomes for adult outpatients with depression under prepaid or fee-for-service financing. Arch Gen Psychiatry 50: 517–525
Roy D (1991) Setting up district audit meetings in psychiatry. Psychiatr Bull 15: 417–418
Royal College of Psychiatrists (1989) Preliminary report on medical audit. Psychiatr Bull 13: 577–580
Royal College of Psychiatrists (1991) Good medical practice in the aftercare of potentially violent or vulnerable patients discharged from in-patient psychiatric treatment. Royal College of Psychiatrists, London
Rutz W, Von Knorring L, Walinder J (1992) Long-term effects of an educational program for general practitioners given by the Swedish Committee for the Prevention and Treatment of Depression. Acta Psychiatr Scand 85: 83–88
Schuster J (1991) Ensuring highest quality care for the cost: coping strategies for mental health providers. Hosp Community Psychiatry 42: 774–776
Shea MT, Elkin I, Imber SD, Sotsky SM et al (1992) Course of depressive symptoms over follow up: findings from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. Arch Gen Psychiatry 49: 782–787
Sheldon T, Borowitz M (1993) Changing the measure of quality in the NHS: from purchasing activity to purchasing protocols. Quality Health Care 2: 149–150
Shepherd G (1988) Evaluation and service planning. In: Lavender A, Holloway F (eds) Community care in practice. Wiley, Chichester, pp 91–114
Sibley JC, Spitzer WO, Rudnick KV, Bell JD, Bethune RD, Sackett DL, Wrigth K (1975) Quality of care appraisal in primary care: a quantitative method. Ann Intern Med 83: 46–52
Stocking B (1992) Promoting change in clinical care. Quality Health Care 1: 56–60
Tarlov AR, Ware JE, Greenfield S, Nelson EC, Perrin E, Zubkoff M (1989) The medical outcomes study. An application of methods for monitoring the results of medical care. JAMA 262: 925–930
Tugwell P (1979) A methodological perspective on process measures of the quality of medical care. Clin Invest Med 2: 113–121
Turner WE (1989) Quality care comparisons in medical/surgical and psychiatric services. Administration Policy Ment Health 17: 79–70
Wells KB, Brook RH (1988) The quality of mental health services: past present and future. Plenum: London
Wells KB, Burnam MA, Rogers W, Hays R, Camp P (1992) The course of depression in adult outpatients. Results from the medical outcomes study. Am J Psychiatry 49: 788–794
Williamson JW (1971) Evaluating quality of patient care. JAMA 218: 564–569
Wing JK (1992) Epidemiologically-based mental health needs assessments. Review of research on psychiatric disorders (ICD-10, F2–F6). Royal College of Psychiatrists Research Unit (Mimeo), London
Wright EC (1993) Non compliance – or how many aunts has Matilda? Lancet 342: 909–913